ESD-TDR-64-523

A GENERALIZED TEACHING MACHINE DECISION STRUCTURE WITH APPLICATION TO SPEED READING

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-523
May 1964

Theodore R. Strollo
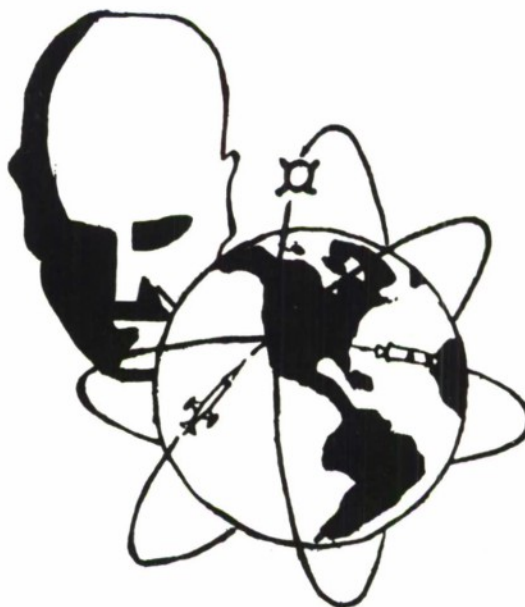
DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 7682
Task 768204

AD602660

Qualified requesters may obtain copies from Defense Documentation Center (DDC). Orders
will be expedited if placed through the librarian or other person designated to request docu-
ments from DDC.

DDC release to OTS is authorized.

2 of 3

TECHNICAL REPORT NO. 7

"A GENERALIZED TEACHING MACHINE DECISION STRUCTURE
WITH APPLICATION TO SPEED READING"
by
Theodore R. Strollo
May, 1964

RESEARCH IN COMPUTER-
AIDED INSTRUCTION

MASSACHUSETTS INSTITUTE
OF
TECHNOLOGY

FOREWORD

The Center for Operations Research at the Massachusetts Institute of Technology is an interdepartmental activity devoted to graduate training and research in the field of operations research. Its products are books, journal articles, detailed reports, and students trained in the theory and practice of operations research.

Philip M. Morse
Director of the Center

George R. Murray, Jr.
Project Supervisor

ACKNOWLEDGEMENT

A GENERALIZED TEACHING MACHINE DECISION STRUCTURE WITH
APPLICATION TO SPEED READING

by

THEODORE ROBERT STROLLO

A relatively new type of automated instruction called the "computer-directed" teaching machine is discussed in this thesis. Typical present-day teaching machines either give every student the same instruction material or choose what material the student receives on the basis of his answer to the last question. The computer-directed machine chooses instruction material by making a statistical evaluation of the student's total behavior in comparison with other students' total behaviors. This machine's statistics are actually changed as new students take the course. Such a teaching machine can perform very much like a human tutor who adjusts his presentation to fit the individual student's capabilities and who improves his teaching technique with each student.

The role of the computer-directed machine in the teaching machine field can only be determined after:

1. A technique for comparing teaching machines is developed.

2. More research is performed utilizing the computer-directed machine.

In this paper a technique is suggested for comparing teaching machines. The machine's tutorial functions would be fitted to a very general model of the tutorial teaching cycle. This allows the various automated instruction devices to be discussed in terms of a common model. An application of the computer-directed machine was made to a speed reading course. Preliminary experiments with this course indicate that the computer-directed machine can perform like a human tutor.

The topic of speed reading lends itself to many possible future experiments. Since most students know something about speed reading prior to the course, the student's speed reading skill before and after the course could be measured and improvements could be noted. Many non-automated courses for speed reading exist, and the students' improvements with automated and non-automated instruction could be compared.

Thesis Supervisor: Richard D. Smallwood
Title: Lecturer, Department of Electrical Engineering

ESD-TDR-64-523

A GENERALIZED TEACHING MACHINE DECISION STRUCTURE
WITH APPLICATION TO SPEED READING

TECHNICAL DOCUMENTARY REPORT NO. ESD-TDR-64-523

May 1964

Theodore R. Strollo

DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Project 7682
Task 768204

FOREWORD

ACKNOWLEDGEMENT

REVIEW AND APPROVAL

This Technical Documentary Report has been reviewed and is approved.

FOR THE COMMANDER

JOSEPH T. BEGLEY   ROY MORGAN
Chief, Applications Division   Colonel, USAF
Decision Sciences Laboratory   Director, Decision Sciences Laboratory

ABSTRACT

A Generalized Teaching Machine Decision Structure with

Application to Speed Reading

A relatively new type of automated instruction called the "computer-directed" teaching machine is discussed. Typical present-day teaching machines either give every student the same instruction material or choose what material the student receives on the basis of his answer to the last question. The computer-directed machine chooses instruction material by making a statistical evaluation of the student's total behavior in comparison with other students.' total behaviors . This machine's statistics are actually changed as new students take the course. Such a teaching machine can perform very much like a human tutor who adjusts his presentation to fit the individual student's capabilities and who improves his teaching technique with each student.

The role of the computer-directed machine in the teaching machine field can only be determined after:

1. A technique for comparing teaching machines is developed.

2. More research is performed utilizing the computer-directed machine.

In this paper a technique is suggested for comparing teaching machines. The machine's tutorial functions would be fitted to a very general model of the tutorial teaching cycle. This allows the various automated instruction devices to be discussed in terms of a common model. An application of the computer-directed machine was made to a speed reading course. Preliminary experiments with this course indicate that the computer-directed machine can perform like a human tutor.

The topic of speed reading lends itself to many possible future experiments. Since most student's know something about speed reading prior to the course, the student's speed reading skill before and after the course could be measured and improvements could be noted. Many non-automated courses for speed reading exist, and the student's improvements with automated and non-automated instruction could be compared.

KEY WORDS

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

THE TEACHING PROCESS

Comparing Instruction Techniques

While research projects on automated instruction are being conducted in many parts of the country, very few attempts have been made to compare the various experiments. (c.f. Skinner, "No large-scale evaluation of machine teaching has yet been attempted. We have so far been concerned mainly with the practical problems in the design and use of machines and with testing and revising simple programs", pg. 159, ref. 17.) This is largely attributed to the lack of a standard notation or measuring stick by which instruction techniques can be compared. If automated instruction devices are ever to become marketable, there must be a way to evaluate them both in terms of other automated devices and conventional instruction techniques. A comparison method would be useful which would answer questions like these:

1. How does this instruction technique accomplish the process of teaching?

2. How is the student paced through the course?

3. How does the structure of the course change after students complete their study?

To facilitate this useful comparison, a model for the teaching process is proposed. Individual teaching techniques could be fitted to this model, and a standard notation would permit techniques to be compared on the basis of how they fit the model. Such a model would be

general enough to cover all variations of the teaching process. Fundamental to the presentation of a model for teaching is an understanding of the mechanics of teaching itself.

Teaching

The goal of teaching is the student's mastery of a topic's principles or skills. A course generally presents the topic principles in the from of sub-topics; thus, the topic is taught in small increments (most researchers, including Skinner[17], in the field of automated instruction agree that optimum learning occurs when the course is composed of a large number of steps with very few sub-topics in each step. This opinion is supported by such experimenters as Coulson and Silberman[5]). A course may be pictured as a series of ascending levels---each level representing a status position in the course indicating that the student who reaches this point has mastered all of the sub-topics marked by the previous levels. When a student reaches the uppermost or final level of the course, he has mastered the whole topic.

Final level of course   level L
                    .
                    .
_____ level 3

_____ level 2

_____ level 1
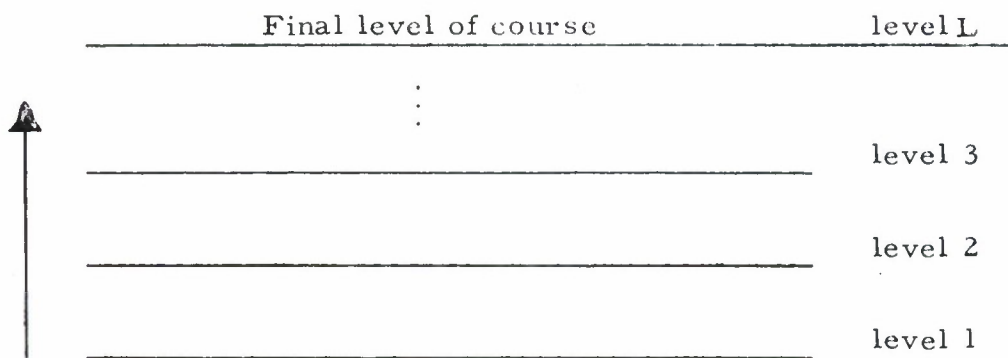
Macro View of a Course

Figure 1-1

Teaching is complicated by the fact that learning is so dependent on the individual. A course presentation that might work very well for one student could be terrible for another. In terms of the macro or over-all view of a course, the teacher's changes in his presentation are revealed by the different paths for each student between levels of the course. The path for a bright student might exhibit skipping over several levels at a time while the path for a relatively dull student might show a tedious level by level ascent.
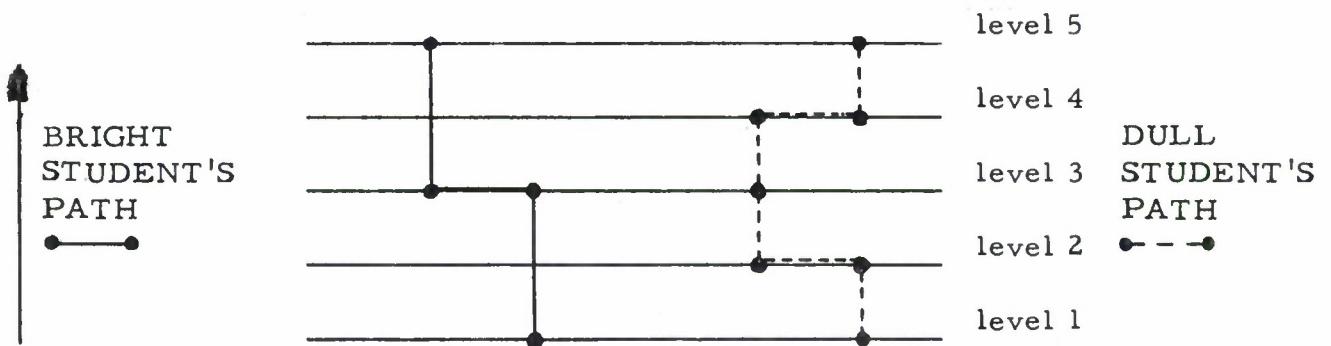
Possible Paths for a Bright and Dull Student

Figure 1-2

The teacher decides how to modify his presentation on the basis of the individual student's learning behavior.

Tutoring is basically a feedback controlled system[15]. The teacher presents material to the input (the student's sensory receptors) attempting to obtain desired responses at the output (the student's test behavior). The responses are analyzed by the teacher who adapts his presentation to get the proper response.

```
┌─────────────────────────────────────────────────┐          ┌──────────────┐
│ TEACHER                                         │          │              │
│ ┌────────┐  ┌─────────────────┐                │  INPUT   │   STUDENT    │
│ │ PRE-   │  │  INSTRUCTION    │───────────────────────────▶│              │
│ │ SENT-  │  │    ELEMENTS     │                │ TO STUDENT│              │
│ │ ATION  │  └─────────────────┘                │          │              │
│ │        │  ┌─────────────────┐                │          │              │
│ │ PLAN   │  │  PRESENTATION   │◀──────────────────────────│              │
│ │        │  │    MODIFIER     │                │  OUTPUT  │              │
│ └────────┘  └─────────────────┘                │FROM STUDENT'S         │
└─────────────────────────────────────────────────┘  TEST   └──────────────┘
                                                    RESPONSES
```
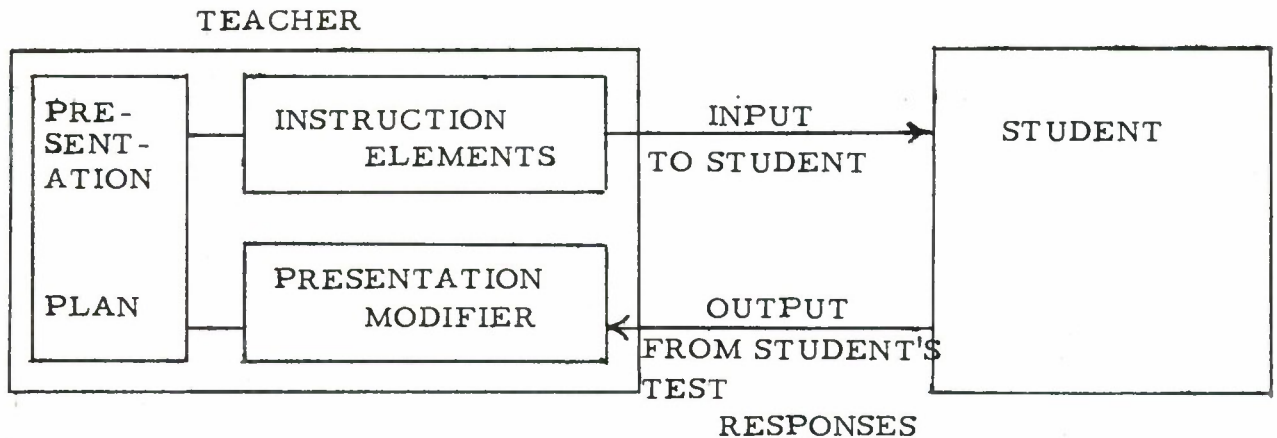
Tutorial Instruction---A Feedback Controlled System

Figure 1-3

One of the tutor's most important functions, then, is the modi-
fication of his presentation. He performs this function by choosing from
his repertoire an appropriate method of instruction. That is, the instructor,
faced with the problem of teaching the course's remaining sub-topics and
having several alternative presentations in his repertoire, chooses the
presentation most suited to his student. He makes this choice periodi-
cally throughout the course because the optimum presentation may change
as the student progresses to new material.

The Teaching Cycle

If a teacher-student environment is observed for some time,
a very definite cyclic behavior is noted. The rhythm of teaching, testing,
and modifying the teaching (based on test results) is plainly apparent.
At first the teacher has some a priori plan of presentation of the material.
Perhaps this plan is based on previous experience with other students,
or it is designed to cover certain material in allotted amounts of time.
The teacher will begin the instruction following this plan. After a while

he will test the student and evaluate the effectiveness of the present plan. The plan is modified to fit the student's needs, and the whole cycle repeats.

1. The teacher chooses the presentation that is best-suited to the student at a given level.

2. The teacher presents this block of instruction.

3. The student is tested on the material covered by this block of instruction.

4. The student is placed at a new level in the course.

The Teaching Cycle (condensed)

Figure 1-4

The teaching cycle can be observed in the macro or overall course model
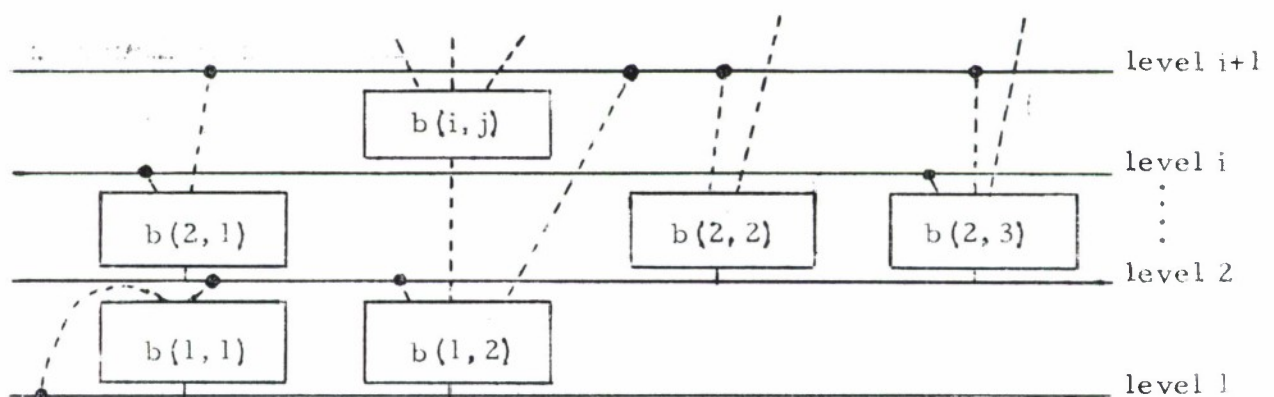
Macro Model Showing Teaching Cycle

Figure 1-5

Each of the methods of instruction available to the teacher at a given level, i, is labeled block $b(i,j)$. The subscript j denotes the particular block or instruction method. Note that at each level, i, the teacher must choose a particular block of instruction $b(i,j)$ appropriate for this student. This is called the tutorial decision making process. After the teacher presents a block of instruction, he tests the student. The student is now placed at a new level in the course because the teacher has re-evaluated the student's mastery of the topic. This is called the tutorial placement function. At this new level, the teacher must again make a decision about a new $b(i,j)$, and the process cycles.

The placement function has been drawn as a quantized function. That is, only a finite number of dotted lines are shown placing each student from level to level via a block of instruction and associated test, yet there are innumerable test behaviors which the student could exhibit. However, there are some very good reasons for quantizing the placement function. The two most significant reasons are:

1. Techniques for measuring learning are, at best, reliable only as discrete measures, not continuous measures, of a student's actual learning. (e.g. placement for all those students with grade "A" behavior might be the same. Similarly placement for students with grade "B", "C",... behavior.)

2. If a course has a finite number of levels, then the number of different placements must be finite.

CHAPTER II

A MODEL FOR THE TEACHING CYCLE

In order to get a more detailed look at the teaching cycle, the macro model of the whole course will be replaced by a micro model of the cycle itself. This micro model has the structure of a tree segment. When a course offers several alternative paths from start to finish, it is described as exhibiting branching. Therefore, a tree is a useful topology for the teaching cycle because it shows the branching nature of a course very adequately.

The levels (i's) of the course are now represented by the level nodes of the tree (the dark circles---see Figure 2-1). The dark branches represent the various blocks of presentation available at each level. The test period is represented by the test nodes (the light circles). The student's test results place him (via the light branches) at a new level node.

The representation of the entire course, by drawing the whole tree with interconnecting micro models, would indeed be more cumbersome than the anlogous representation by the macro model. However, the representation of the teaching cycle alone may now be considered in minute detail.

The micro model is drawn to show the nth teaching cycle in the course (i.e. the next cycle after n-1 previous cycles have been executed). Thus i_n is the present status level of the student, and the teacher must choose an instruction block or branch from the available values of j_n (the tutorial decision making process). The student is tested after terminating the branch, j_n. His test behavior will lie in one of the k_n discrete ranges and will place him at a new level i_{n+1} (the tutorial placement function).

-7-
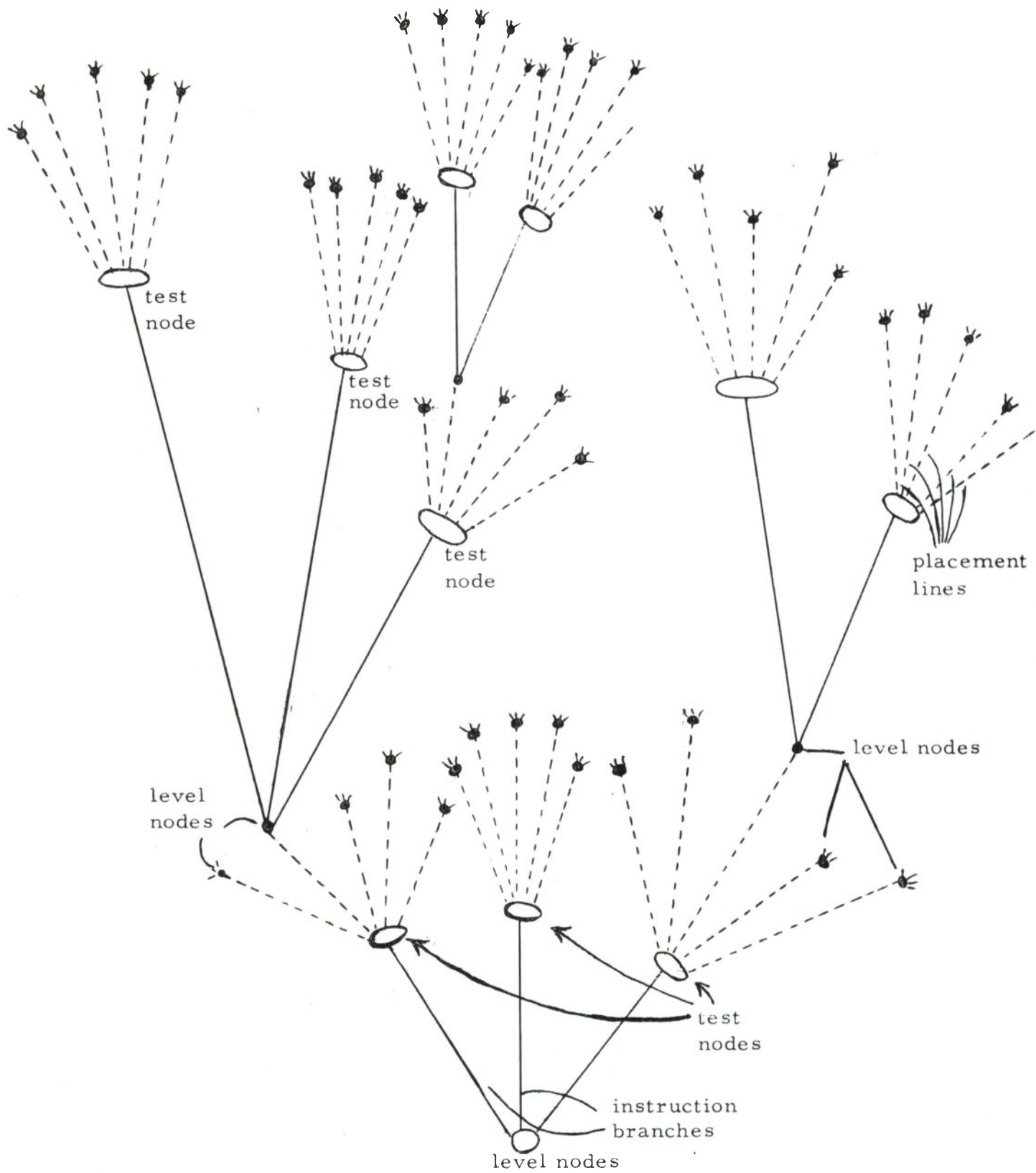
Tree Model of a Segment of
The Course

Figure 2-1
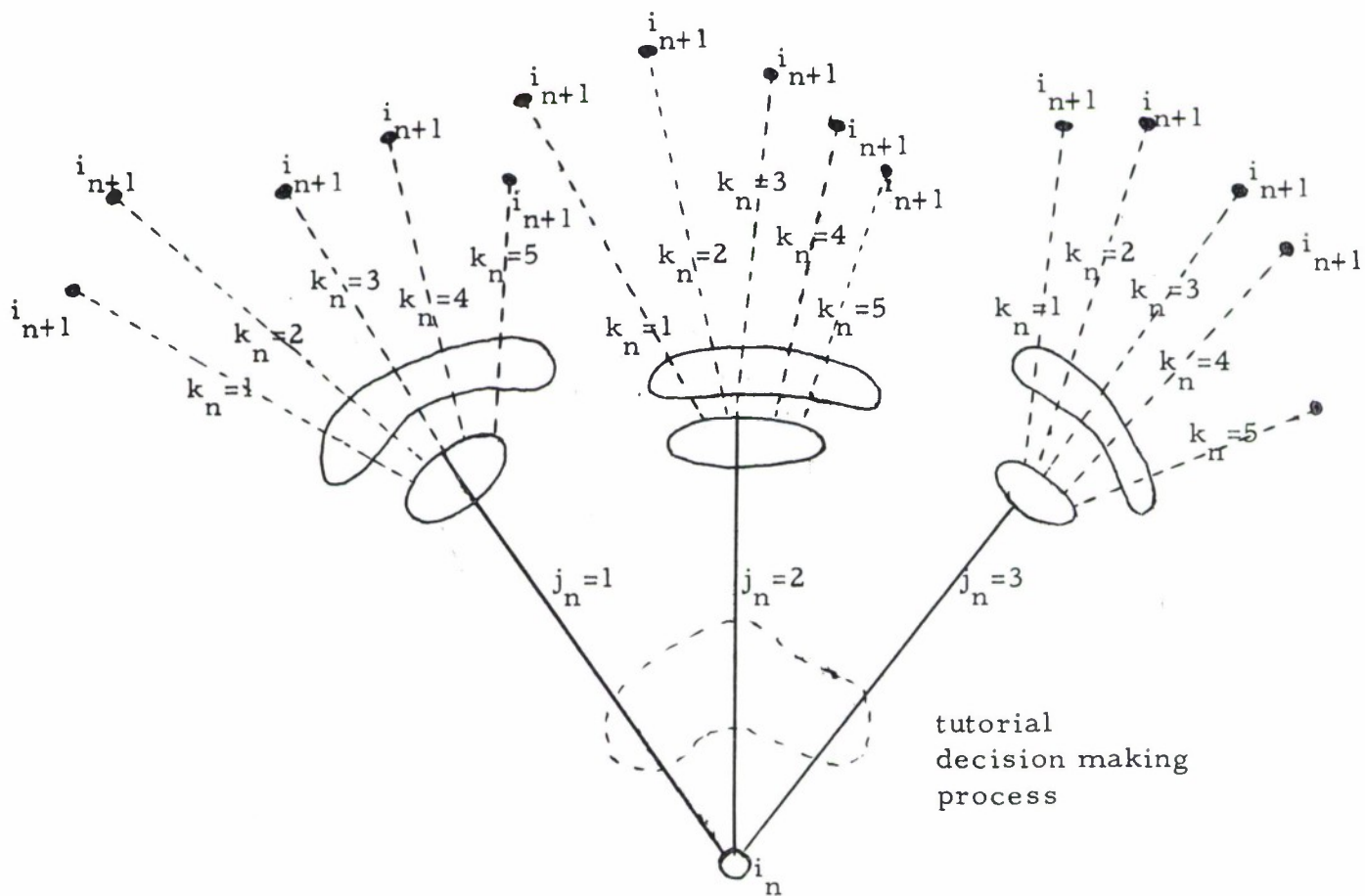
Micro Model of the Teaching Cycle

Figure 2-2

CHAPTER III

FITTING CURRENT EXPERIMENTS TO THE MICRO MODEL

Straight Line or Linear Teaching

By far the most common method of teaching used today is the lecture method which employs straight line or linear teaching. That is, there is simply no branching; the students follow a pre-selected path through the course. This method of teaching fits the model by reducing it to a trivial form. One can think of each level's having only one branch (j_n) and each test behavior (k_n) leading to the same placement (i_{n+1}). Also one can suppose, since relatively few tests are given during the course (they are no longer needed to guide the teacher in his path determination---merely to grade the student), that the individual blocks of instruction, $b(i_n, j_n)$, become longer and the total number of teaching cycles in the course becomes reduced (the step-size increases).

Intrinsic Programming

Crowder defines a method of course design called intrinsic programming[8]. The choice of the proper alternative instruction block is built into the instruction material itself; so that, the material may be self-taught. An example of an intrinsically programmed device is the so-called programmed text which is well represented by the Crowder Scrambled Book[8]. With these texts, the choice of the next page to be read is determined by the student's answers to questions on the present page; the choice is independent of the answers to previous questions.

Of course, on each page there is only one mode or block of instruction available. This description applies to a number of auto-instructional devices currently on the market (such as Auto-Tutor[19]).

This type of instruction fits the model very well. Leaving each level (i_n) there is still but one branch (j_n), but now the test behavior ranges (k_n) are definitely used to place the student at the next level (i_{n+1}) or page.
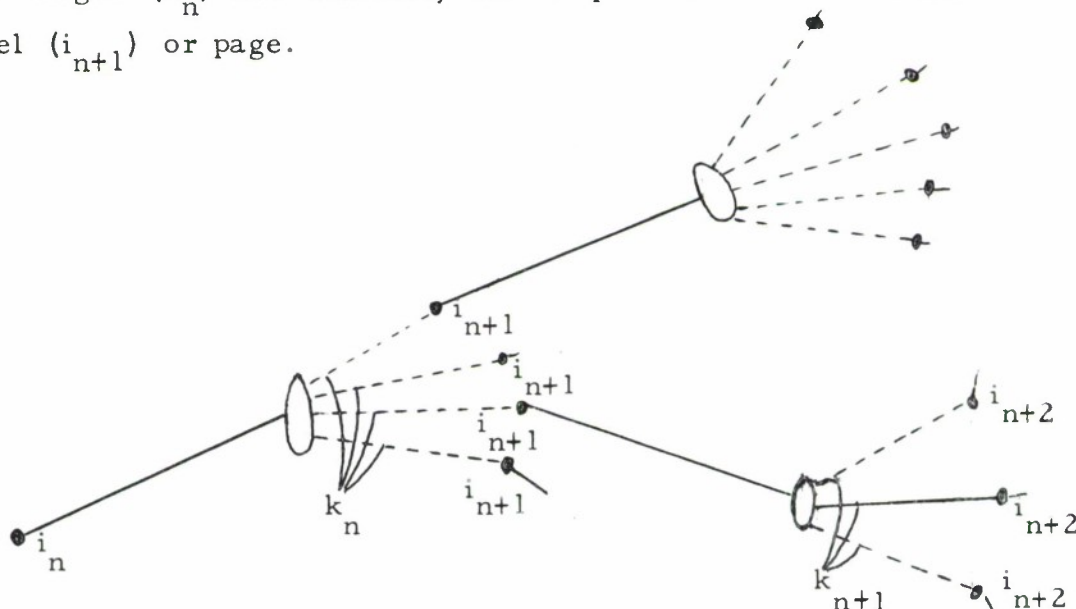
Model for a Programmed Text

Figure 3-1

Extrinsic Programming

Crowder defines extrinsically programmed courses[9] as those where the choice of alternative instruction blocks (branching) is performed by an external element such as a teacher or a computer, and the basis for this choice involves the student's cumulative test behavior. A typical computer-based teaching machine is a facility

called "CLASS" developed by the Systems Development Corporation of Santa Monica, California[16]. Many of the courses taught at "CLASS" may be described by the following structure:

Courses are split into sub-topics A, B, C, D, ... Available at each level are several alternative instruction modes. The alternatives are organized such that alternative I covers topic A in just a brief manner, alternative II goes into more detail, alternative III goes into still more detail, etc.
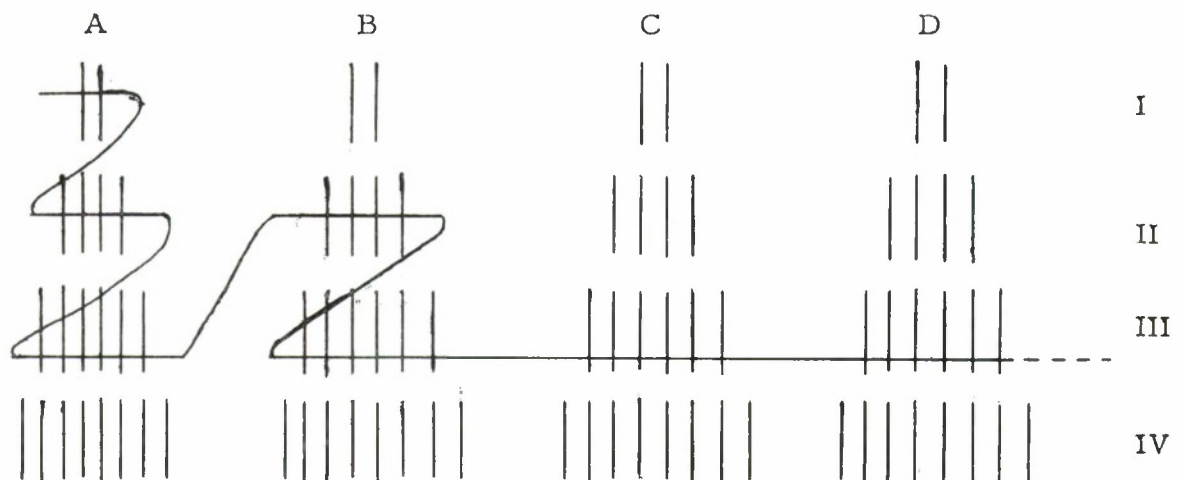
Organization of a Computer-Based Course

Figure 3-2

A student might be initiated from topic A on alternative I. If he does not perform well, he may be routed through alternative II. Suppose he is also routed through alternative III before he masters topic A. Now the computer decides that this student is not as bright as at first anticipated, and he perhaps needs a more detailed coverage of future topics. Therefore for topic B, he might be initiated on alternative II. Suppose he drops again to alternative III. Next the computer would initiate him on topic C alternative III, etc.

In applying this type of automated instruction to the model, the level placements based on the student's test behavior are similar to the intrinsic Crowder type, but the choice of the instruction alternative is decidely extrinsic. With this type of teaching machine, the full scope of the teaching cycle model is represented.

Skinner Disc Device

The Skinner Disc Device[17, 19] is a very difficult one to fit to the micro model of the teaching cycle. The Skinner Disc or Tape presents material to the student in an order identical to the physical sequence on the disc or tape. With each frame of material there is a question. When the student answers the question correctly, the frame is dropped out of the course material. The disc or tape is rerun until all frames are dropped out, and theoretically, all of the material is learned.

This fits the model if one is willing to accept the idea of disappearing branches. Possibly this can be represented if one considers each re-show of the tape or disc as a new part of the course and not just a re-traverse of the course.

Of course no model can be expected to represent adequately all of the specific cases which it generalizes. The Skinner Disc Device is a very unusual teaching method, and most present day teaching methods are more like the previously described teaching techniques. However the model represents a large percentage of present-day teaching situations.

13-

CHAPTER IV

COMPUTER-DIRECTED TEACHING MACHINES

The decision functions performed by most present-day computer-based teaching machines are intuitive and somewhat arbitrary judgements. With most of these machines the behavior of each student is forgotten as the next student is encountered. Smallwood[18] envisioned a teaching environment as a probabilistic system---a system in which decisions are based on statistical comparisons of the present student's behavior with previous student's behaviors. Such a teaching system would continuously revise its statistics about past students as new students took the course. Smallwood constructed a computer simulation of this teaching system. We call this type of teaching machine the "computer-directed" teaching machine.

The author of this paper is presenting another computer-directed teaching machine utilizing a probabilistic decision structure. Now the notation has been established, and a modified decision structure which can apply to many courses has been developed. The computer-directed decision mechanism will be presented in this section.

The tutorial processes which tailor the course to an individual are two-fold:

1. The decision making process which chooses an instruction block from a number of alternatives.

2. The placement function which re-evaluates the student's mastery of the topic by placing him at an appropriate level in the course.

Computer-directed machines are distinguished from computer-based machines by the different realizations of the tutorial processes.

Computer-Directed Realization of the Placement Function

Placement is the process of assigning a student to a new level after re-evaluating his mastery of the topic. Therefore placement is a function of the student's old level (i_n), the instruction branch (j_n) which the student was given at the old level, and the range (k_n) into which the student's branch test behavior fell. With this computer-directed teaching machine the placement function is pre-determined by the structure of the course. For example, if the student were initiated on a branch which covered several sub-topics and if he did very well on the branch test, he would probably be skipped ahead a couple of levels. Whereas, if the student were on a branch covering only a few sub-topics and if he did well, he would probably just be advanced to the next level.

Expressing this placement function mathematically,

$$i_{n+1} = V(i_n, j_n, k_n) \tag{4.1}$$

This function remains constant as the course is taught to successive students.

Ideally it might be desirable to change the placement function as well as the course structure by some course monitor that observes the reactions of students to the present structure. Such a course monitor or automatic course programmer, while beyond the scope of this project, is worthy of consideration.

Computer-Directed Realization of the Decision Process

The decision process chooses the instruction block best-suited to the student from the entire repertoire of alternate instruction blocks available at the present level. Differences between decision processes result from different interpretations of the words "best-suited".

In this research the words "best-suited" were interpreted to mean the choice of that instruction block that will maximize the expected value of a parameter indicative of the student's mastery of the topic. This is a reasonable interpretation because the goal of a course is to enable the student to master the topic.

Let this parameter be called U which represents the student's learning or mastery of the course material. Also let h_n represent the student's cumulative past history (test behavior) generated before the nth teaching cycle. Then it is desired to find that branch (j_n) leaving the present level (i_n) which maximizes the expected value of U given h_n.

$$\underset{j_n}{\text{Max}} \ \overline{U}_{i_n j_n}(h_n) \text{ is desired.}$$

The notation $\underset{j_n}{\text{Max}}$ means that j_n for which the function $\overline{U}_{i_n j_n}(h_n)$ is a maximum.

This value may be expressed formally from the expected value theorem[21] as:

$$\underset{j_n}{\text{Max}} \ \overline{U}_{i_n j_n}(h_n) = \underset{j_n}{\text{Max}} \int\limits_{\text{over } U} U \ f_{i_n j_n}(U|h_n) \, dU \qquad (4.2)$$

In order to evaluate this expression it is necessary to express the conditional probability density function $f_{i_n j_n}(U|h_n)$ in terms of statistics which are easily derived from students' path and test behaviors. Consider that the cumulative history at the beginning of the next cycle, h_{n+1}, is a function, W, of the old cumulative history, h_n, and the test behavior range for this cycle, k_n. With the present teaching machine structure the cumulative history is simply a uniformly weighted average of all of the student's test behaviors.

$$h_{n+1} = \frac{nh_n + U(k_n)}{n+1}$$

(4.3)

where $U(k_n)$ is the value of the parameter U before it is fitted to range k_n. (i.e. the actual history generated by the student during the present cycle.)

While it may be argued that a great deal of information is lost about the student's behavior during each cycle by uniformly averaging his behaviors, it is important to simplify the representation of the student's history to a single parameter (such as cumulative or averaged histories) because of the large number of calculations involved in choosing the appropriate instruction block. Now we consider all possible ranges the student's test behavior might fit for the block $b(i_n, j_n)$. The function $f_{i_n j_n}(U|h_n)$ (abbreviated f_n) can be expressed as the sum of the probability of each test behavior range times the function $f_{i_{n+1} j_{n+1}}(U|h_{n+1})$ (abbreviated f_{n+1}) for all possible ranges.

$$f_{i_n j_n}(U|h_n) = \sum_{k_n} P_{i_n j_n}(k_n|h_n)\, f_{i_{n+1} j_{n+1}}(U|h_{n+1})$$

(4.4)

where $i_{n+1} = V(i_n, j_n, k_n)$

$h_{n+1} = W(h_n, k_n)$

-17-

It is possible to evaluate $P_{i_n j_n}(k_n | h_n)$ (the probability of each test behavior range conditioned upon a given past history) easily in terms of student path and test behaviors. From Bayes Theorem[21] the probability of a particular behavior range given a certain past history is equal to the conditional probability density function for this past history (given the student's test behavior fitted the specified range), times the probability that the student's behavior will lie in this range, divided by the probability that the student had this past history.

$$P_{i_n j_n}(k_n | h_n) = \frac{g_{i_n j_n}(h_n | k_n) \; p_{i_n j_n}(k_n)}{\sum_{k_n} \text{numerator}} \tag{4.5}$$

The probability, $p_{i_n j_n}(k_n)$ (abbreviated p_n), is estimated by that fraction of the number of students reaching level i_n and emerging on branch j_n whose test behavior falls in range k_n. The conditional density function $g_{i_n j_n}(h_n | k_n)$ is estimated by observing the past histories of those students who reach level i_n, emerge on branch j_n, and whose test behavior lies in range k_n. A density function (for the present machine Beta functions are used---see Appendices C and D) is fitted to these observations[7,18]. (Note that Smallwood's[18] decision structure determines $P_{i_n j_n}(k_n | h_n)$ in terms of an intuitive probability model which, while reducing calculation time, is not as mathematically justifiable as the Bayes Theorem expansion and subsequent estimation.)

The expression of f_n in terms of f_{n+1} can be extended successively to $f_{n+2}, \ldots$ until the last level, i_ℓ, (the end of the course hence $i_\ell = L$, see page 2).

$$f_{i_n j_n}(U|h_n) = \sum_{k_n} P_{i_n j_n}(k_n|h_n) \sum_{k_{n+1}} P_{i_{n+1} j_{n+1}}(k_{n+1}|h_{n+1}) \ldots f_{i_\ell}(U|h_\ell)$$

$$(4.6)$$

Thus the mean value of U for a student at the $i_n{}^{th}$ level can be maximized by picking that j_n for which

$$\sum_{k_n} P_{i_n j_n}(k_n|h_n) \underset{j_{n+1}}{\text{Max}} \sum_{k_{n+1}} P_{i_{n+1} j_{n+1}}(k_{n+1}|h_{n+1}) \ldots \int_{\text{over } U} U f_{i_\ell}(U|h_\ell)\, dU$$

$$(4.7)$$

is a maximum.

      At the final level of the course there is only one instruction block after which the final test is given. Therefore $\underset{j_\ell}{\text{Max}}$ is meaningless since there is only one j_ℓ which is therefore the maximum. The only quantity in the maximization expression which remains to be discussed is the integral:

$$\int_{\text{over } U} U f_{i_\ell}(U|h_\ell)\, dU$$

$$(4.8)$$

This integral may be approximated by the sum:

$$\sum_{k_\ell} \tilde{U}(k_\ell) P_{i_\ell}(k_\ell|h_\ell)$$

$$(4.9)$$

where $\tilde{U}(k_\ell)$ is the average value of U in the k_ℓ range.

It is now obvious that $P_{i_\ell}(k_\ell|h_\ell)$ can be estimated in the same manner that $P_{i_n j_n}(k_n|h_n)$ is estimated.

Notice that in order to determine $\underset{j_n}{\text{Max}}(f_n)$, one must know $\underset{j_{n+1}}{\text{Max}}(f_{n+1})$, but to determine $\underset{j_{n+1}}{\text{Max}}(f_{n+1})$, one must know $\underset{j_{n+2}}{\text{Max}}(f_{n+2}), \ldots$ It becomes apparent that one must know $\underset{j_{\ell-1}}{\text{Max}}(f_{\ell-1})$ before determining $\underset{j_{\ell-2}}{\text{Max}}(f_{\ell-2})$ and so on. This suggests a dynamic programming technique for computing $\underset{j_n}{\text{Max}}(f_n)$. First an initial path from level i_n (starting with the first possible j_n) to the end of the course is routed. Working __backwards__ from the test level, $\underset{j_{\ell-1}}{\text{Max}}(f_{\ell-1})$, $\underset{j_{\ell-2}}{\text{Max}}(f_{\ell-2}), \ldots, \underset{j_{n+1}}{\text{Max}}(f_{n+1})$ are found, and the expected value of the parameter U for this first value of j_n is determined. This process is repeated for all possible j_n, and the j_n giving the maximum value for the parameter U is chosen.

Tree structures of a course become extremely complicated, and, by nature of the branching, get increasingly complicated with each cycle beyond the present level node. In fact the tree can become infinite. Increasing complication means increasing computation time. In order to reduce computation time, it is desirable to truncate the exhaustive search evaluation of $\overline{U}_{i_n j_n}(h_n)$ before reaching the test level. That is, suppose the search is truncated after nmax future teaching cycles are spanned. This can be done if one is willing to estimate h_ℓ in terms of h_{n+nmax}. If one uses this approximation, it is not necessary to determine the P_m, where $m > n+nmax$, because the sum of these P_m's over all paths leading from node m to the end of the course is simply equal to unity. (All paths emerging from node m eventually lead to the end of the course.) This truncation strategy leads to a modified decision formula:

let $M = n+nmax$

$$\underset{j_n}{\text{Max}}\ \overline{U}_{i_n j_n}(h_n) = \underset{j_n}{\text{Max}} \sum_{k_n} P_{i_n j_n}(k_n | h_n) \ldots \underset{j_M}{\text{Max}} \sum_{k_M} P_{i_M j_M}(k_M | h_M) \int_{\text{over } U} U f_{i_\ell}(U | \hat{h}_\ell)\, dU$$

where $\hat{h}_\ell$ is the estimate of h_ℓ \hfill (4.10)

The method used for estimating h_ℓ by this machine was to consider:

$$\hat{h}_\ell = \text{function } (h_M, i_\ell - i_M) \tag{4.11}$$

The average change in history per single level advance is measured--- call it Δh_{ave} .

Then

$$\hat{h}_\ell = h_M + \Delta h_{ave}(i_\ell - i_M) \tag{4.12}$$

Smallwood's machine and the present machine utilize the truncation strategy in the decision technique. However, Smallwood's machine was limited to a fixed, 3-step future search while the present machine is capable of an arbitrary number of step future search (including an exhaustive search to the end of the course---see Appendix B).

CHAPTER V

PROGRAMMING A COURSE
FOR THE COMPUTER-DIRECTED MACHINE

Programming or planning a course[2, 14, 19] for the computer-
directed teaching machine involves several systematic operations:

1. The course must be divided into sub-topics and consequently
into levels.

2. At each level a number of different instruction blocks for
presenting the new sub-topics must be constructed. Some of the blocks
should present several sub-topics (with the goal of multi-level skips for
fast students) while other blocks should present just a few sub-topics
(with the goal of bringing the student to the next sequential level). While
some of the blocks should give a concise presentation of the material,
others should supplement the material with examples.

3. The course designer must tabulate a placement function
$V(i, j, k)$ which reflects his opinions about where a student should be
placed given his present placement level (i) , present block of instruc-
tion (j) , and present test behavior (k) .

4. A table of <u>a priori</u> estimates for the $p_{ij}(k)$ and $g_{ij}(h|k)$
functions must be constructed.

5. While multiple parameters are generally good indications
of the student's achievement in the course, these parameters must
be expressed in terms of a common parameter U . That is, though

there may be many parameters that indicate the student's mastery of the topic, the decision structure is set up to choose the optimum presentation by maximizing the expected value of the single parameter U. For most topics the multiple parameters can be expressed in terms of a single parameter, but this is not true for all topics.

Suppose the student's achievement in the course is related to the measurement of two parameters --- $u1$ and $u2$. While it is desirable that the measures of $u1$ and $u2$ be individually large, a relatively small measure of $u1$ could be tolerated in association with a relatively large measure of $u2$, and vice versa. This idea is illustrated by the tradeoff curve below.

Utility Tradeoff
Figure 5-1
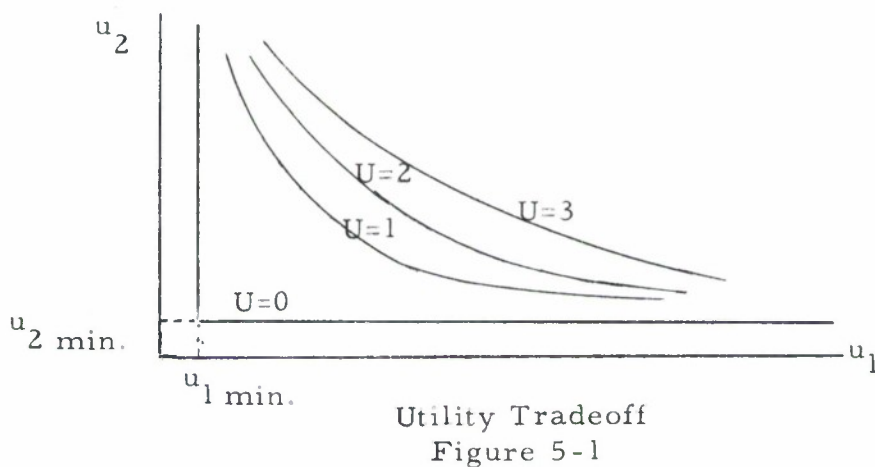
Any point on the same U indifference contour (also called equal-utility contours) is considered to represent the same degree of mastery of the topic.

6. The course should be presented to many test students who are routed through pre-determined paths in the course. These paths are determined so that statistics can be gathered about every block of instruction at each level. These statistics are then used to update the priors of operation 4 (on the previous page).

CHAPTER VI

SPEED READING

Speed Reading as a Topic for a Computer-Directed Machine

The topic chosen for this teaching machine is speed reading. Speed reading was chosen because:

1. One of the most difficult tasks for a teaching machine is the measurement of the student's learning. However the reading rate, which is certainly an indication of a student's mastery of speed reading, is easily measured.

2. Most teaching machines teach a topic new to the student. Everyone taking the speed reading course must, as a prerequisite, know how to read; so, a speed reading course attempts to improve a skill that the student already possesses. The student's speed reading skill could therefore be evaluated before and after the course, and the student's improvement would be easily determined.

3. The input-output devices available for this project are particularly suited to speed reading (see Section VII and Appendix A).

Teaching Speed Reading

Reading is of vital significance to modern man who must be well-informed of the fast-moving stream of events in this modern world. Reading helps to keep him informed. However, the bulk of material he must read is ever increasing, and man must increase his reading speed to meet the new demand. His comprehension of the material he reads must not suffer from his increased speed.

Speed reading courses focus, then, on two objectives:

1. An attempt to increase the reading speed by increasing the efficiency of eye movements and introducing phrase reading techniques.

2. The formation of good reading habits directed towards increased reading comprehension.

Increasing the Reading Speed

Physiologists and psychologists have determined that readers do not scan lines of text smoothly as they read[1]. Indeed the human eye cannot perceive detail while it is in motion. Instead readers scan text with a repeated pattern of sweeps and fixations. It is during the fixation time, which is about 10 times longer than the sweep time[1], that the words are actually read. Rapid readers keep the number of fixations per line to a minimum (about 3 or 4 fixations in a 12 word line[4]).

In order to reduce the number of fixations, one must view more material during each fixation. Efficient readers do not read letter-by-letter but read by phrases or groups of words. Cole[4] describes the progression of a student to phrase reading as a four-stage process:

1. Beginning readers learn to read at first by word spellings---an alphabetical or letter-by-letter approach.

2. Later students read phonetically (which is often a difficult task with the English language).

3. When the readers become familiar with a set of words, they use the look-and-say approach recognizing whole words as they read.

4. The good reader ultimately recognizes whole phrases (or groups of 3 or 4 words) in context.

Several methods are used to train students for phrase reading. Cole[4] suggests the use of texts where the material is organized into phrase groupings which are spaced apart.

e.g.

In the meadow   the brown cows   graze frequently

Students would therefore be forced to read groups of words. The spaces are gradually decreased on successive pages of the text. As the student is conditioned to phrase reading, the original stimulus is vanished and is replaced by the student's own ideas about phrase groupings.

Another technique utilizing a tachistoscope gives students practice with phrase reading. The tachistoscope is a device which presents material for a brief amount of time (1/10 to 1/200 second typically). It was developed during World War II to train pilots[3] to recognize military objectives in a single glance. Students are first shown material for long time periods (1/10 second) then the amount of material increases while the time period decreases (1/200 second). The tachistoscope forces the student to increase the amount of material he can read in a single glance.

The type of material flashed by the tachistoscope is closely related to the student's accuracy and limiting speed of observation. Gray states[20], "Tachistoscopic studies show words whose meanings are familiar are recognized far more rapidly and accurately than nonsense syllables (and numbers)". Therefore a student who performs well on tachistoscopic exercises containing nonsense syllables or numbers must be considered more advanced than the student who performs well with sense words and phrases.

Improving Reading Comprehension

Good reading comprehension is a skill based upon good reading habits. The good reader[1,20]:

1. Concentrates on what he is reading by choosing his study area carefully to minimize distractions.

2. Always improves his vocabulary by looking up words which are new to him.

3. Is able to identify the author's purpose or viewpoint.

4. Is able to note detail and to discriminate.

5. Can concisely and accurately summarize an article.

Habits, good or bad, take considerable time to develop. Good reading habits, being no exception to the rule, take years to develop and must always be maintained. One can make important strides towards improving his reading habits by always being conscious of good reading techniques.

Speed reading courses generally start students off on the road to developing good reading habits by making the student aware of good reading techniques. The student is taught to recognize the author's sign-post---the titles, sub-titles, and topic sentences the author uses to summarize his own material. Students are taught to discriminate significant from insignificant detail.

A typical method for presenting a comprehension improvement course involves the student's reading quantities of text then answering questions about the text that evaluate his reading habits.

Often the text includes hints about reading efficiently, but the text material is usually varied; so that, the student will develop good reading habits for all types of material.

Of the two parameters, good comprehension is generally considered to be of higher value than rapid reading rate. The reader who sacrifices all understanding to reading large volumes of material accomplishes little or nothing at all. Good speed reading courses repeatedly emphasize the importance of thorough comprehension.

CHAPTER VII

APPLICATION OF THE COMPUTER-DIRECTED TEACHING MACHINE
TO A SPEED READING COURSE

Programming the Speed Reading Course

The speed reading course programmed for the computer-directed teaching machine is divided into two parts---part one utilizes a unique tachistoscope (see Appendix A) to train the student for phrase reading. Part two presents ideas about good reading habits directed toward improving comprehension and increasing reading speed.

The course was designed in the manner described by Chapter V. First each part of the course was divided into sub-topics or skills and the levels were assigned. In the tachistoscopic training portion of the course there are really no sub-topics since a single skill is being developed--- phrase reading. Here the increasing levels were assigned to increasing flash rates (decreasing flash duration time) of the tachistoscopic material. The tachistoscope portion was divided into five separate levels. Level 1 corresponds to a flash duration of 1/10 second while level 5 corresponds to a flash duration of 1/100 second with material approximately ten times more complex than that in level 1.

At each of the levels of the course it was necessary to make available several alternate instruction blocks. The tachistoscope alternatives were sense material, nonsense material, and mixed sense-nonsense material in order of increasing difficulty. For the sake of organization, this same order was used in the assignment of the instruction block numbers (j's).

That is $j=1$ or 2 corresponds to sense material, $j=3$ or 4 to nonsense material, and $j=5$ or 6 to mixed sense-nonsense material.

Since the mixed sense-nonsense syllables are part one's most difficult tachistoscopic material, one would expect the placement function to exhibit longer level skips for excellent test behavior with the higher valued j's than for similar test behavior in the lower valued j's. Indeed the construction of the placement function for the speed reading course was based on just this type of reasoning.

The tachistoscopic phrase material was taken from the Phraseoscope[3] slides of the Encyclopaedia Brittanica's Better Reading Program. This material progresses from simple one word flashes (at the lowest levels) to 8 or 9 word phrases and 9 digit number flashes (at the highest levels).
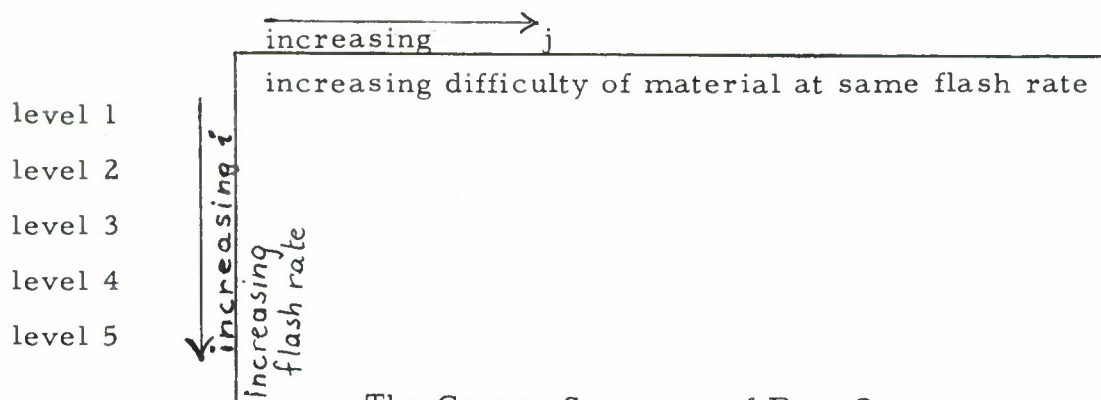
The Course Structure of Part One

Figure 7-1

The second part of the course was easily divided into five sub-topics.

Level 6 concerns the importance of reading accurately and rapidly.

Level 7 discusses the sweep-fixation nature of eye motion during reading and suggests reducing the number of fixations to a minimum.

Level 8 concerns the art of skimming a text by looking for the author's sign posts.

Level 9 is about concentrating while one reads.

Level 10 discusses the art of comprehending material, discriminating important from unimportant details, and finding the main ideas in an article.

The material for part two was taken from "Reading Skills"[1] distributed by the Encyclopedia Brittanica's Better Reading Program[3] and from "Reading Critically"[12]. Ideas are presented about good reading habits in large blocks of text; then the student is asked questions about the text.

Again it was essential to offer several alternate blocks of instruction at each level. Here the low numbered blocks (low value of j) contain large quantities of very simple text about good reading habits with many questions about the text. The high numbered blocks (high value of j) generally contain less material than the low numbered blocks. This material consists of short summaries of the ideas presented in the low numbered blocks, examples of text on which the student can practice the techniques suggested in the summary, and fewer but more difficult questions about both the summary and example texts.

Notice that the organization of part two is consistent with the organization of part one. The part two alternatives become increasingly more difficult (not only are the questions harder, but the student is expected to learn the same sub-topic with less instruction), as the

instruction block number increases.

increasing j →

level 6 | increasingly difficult presentations of same sub-topic
level 7
level 8
level 9
level 10

increasing i ↓ different sub-topics
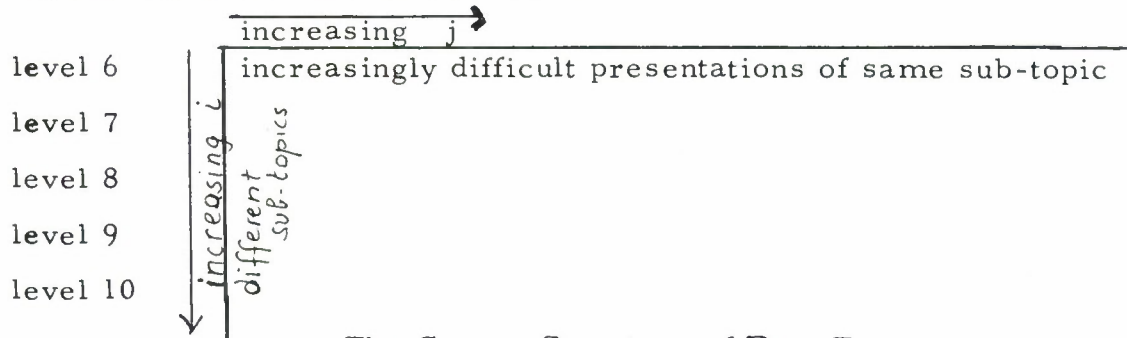
The Course Structure of Part Two

Figure 7-2

The placement function again exhibits longer level skips for students with excellent text behaviors who were instructed by the higher numbered blocks than for those who were instructed by the lower numbered blocks.

The next step of the programming calls for the estimation of the priors for the pertinent probability functions. Since little was known about the reactions of students to this completely new course, the author chose uniform priors---probability functions which predict all paths through the course are equally likely to produce optimum learning.

The Physical Teaching Machine

The actual speed reading course was coded for the dual PDP-1 (Digital Equipment Corporation's Programmed Data Processor - 1) computer system at the Air Force Cambridge Research Laboratory of L. G. Hanscom Field. The tutorial functions were performed by one computer (computer A) while the input-output devices, which presented the actual course, were controlled by another computer

(computer B). The primary reason for using two computers was that the decision calculations took so much time that it was decided to perform these calculations on one machine <u>while</u> the student was taking the course presented by the other machine (see Appendix B).

All course instructions, text material, and questions were presented page by page on the text display scope (see Appendix A). The student would read this material and then proceed to the next page or next part of the course by activating the page turning switch. Questions presented on this scope were answered on the computer typewriter. The student's response to a question was always reinforced by comments on the text display indicating if the student was right or wrong. If he was wrong, the correct answer was indicated.

The tachistoscopic material was presented on the tachistoscope display (see Appendix A). This material was flashed every time the student pressed the flash button. The student was expected to typewrite the syllables or numbers he observed. Again the student's response was reinforced. If the student was correct, he would proceed to new material. If he was incorrect, he had the option of trying again (reflashing and retyping) or giving up (by engaging the give-up switch). The student who gave up was given some partial credit for his last typewritten response. The maximum possible score for the student who tried again was reduced in proportion to the number of times he attempted to read the same material.

Interpreting the Student's Test Behavior

According to Article 5 of Chapter V it is necessary to express all of the parameters indicative of the student's learning in terms of a single parameter U in order to meet the present decision

structure's requirements. In a speed reading course there are two parameters that are pertinent to the student's speed reading skills---his reading rate and his comprehension test scores.

Most speed reading authorities[1, 3, 4, 20] agree that the reader's comprehension is far more important than his reading speed. Therefore comprehension was weighted heavily in expressing the two parameters, reading speed and comprehension, in terms of a single, aggregate parameter U . The simple function chosen to combine the two parameters in this experiment was:

$$U = C^2 S$$

(7.1)

where C is the comprehension test score
       S is the reading rate in words per minute

This function has three desirable characteristics:

1. Comprehension is weighted more heavily than reading speed.

2. Because of the concave nature of the isoquants of U , when the value of one parameter is small while the other is large, the increase in U is much more for a given increase in the small valued parameter than for the same increase in the large valued parameter. Therefore readers who either comprehend very well but read slowly or read rapidly but comprehend very little are given much more credit for improving their deficient skill than their proficient skill. This philosophy discourages "one-sided" readers.

3. The parameter U is easily calculated with this function.
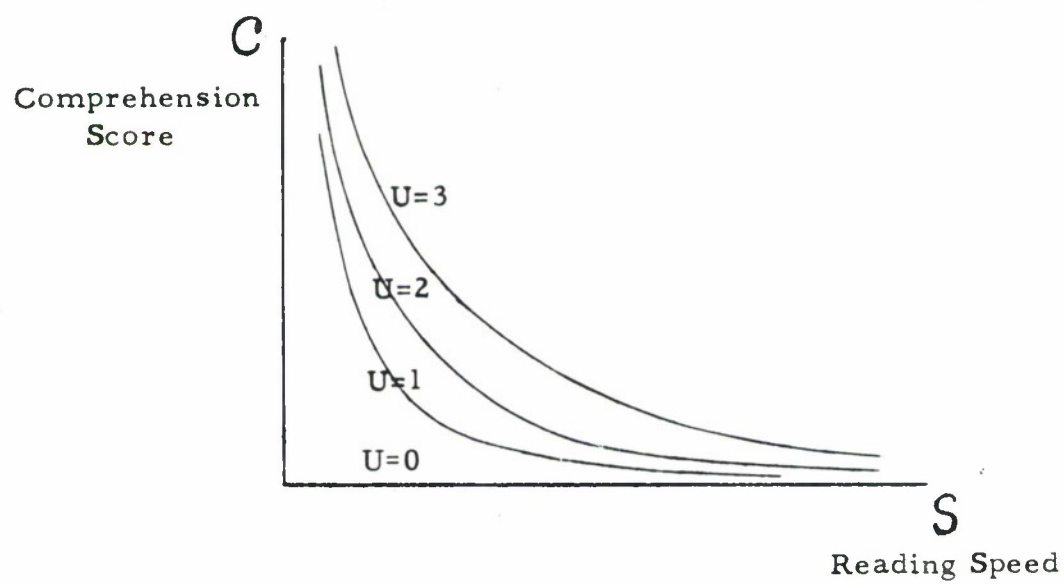
Graph of the U Function Used in This Study

Figure 7-3

CHAPTER VIII

RESULTS, CONCLUSIONS, PRACTICAL CONSIDERATIONS

Many experiments can be performed with the speed reading teaching machine. In this project an investigation of the comparison between students who were taught speed reading linearly versus computer-directed was performed. Some of the other types of experiments that could be performed with the speed reading automated course are suggested in the next section.

A student receiving the linear version of the course would be routed through every possible instruction block $b(i, j)$ at each level i. The student would require about five hours to complete the course for this path. Average students who receive the computer-directed course would be routed through only about 1/3 of the course material and would require no more than two and a half hours to complete the course. Experiments are currently being performed to determine whether both students become equally skillful in their speed reading techniques.

At the present time seven students have taken the speed reading course. They have all been given linear versions of the course. Each student has been routed through a new path in order to accumulate data about as many instruction blocks as possible. This data is currently being used to update the course statistics. Table 8-1 shows the path and test behavior for a typical experiment student. The maximum value of the parameter U has been normalized to 1.0 in all of the following tables.

LEVEL (i)	BLOCK (j)	BEHAVIOR (U)
1	1	.78
3	2	.71
5	3	.86
6	1	.57
7	1	.40
8	1	.62
8	2	.73
9	1	.83
10	1	1.00
11 (the final test)	1	.99

Typical Path and Behavior for Experimental Student

Figure 8-1

After considering table (8-2), the manner in which the computer-directed teaching machine modifies its presentation to suit the individual is revealed. The table shows the computer decision for the appropriate instruction block $(b(i, j))$ at each level (i), the student's test behavior (k) for this block, and the student's placement level based on the previous levle (i), the instruction block $(b(i, j))$, and the test behavior (k). The paths for three hypothetical students with different speed reading abilities are shown. (The current actual course statistics were used for these hypothetical students.) Note that for an excellent student, the student would be progressed rapidly through the course (he would make several multi-level skips), and the student would be given the more difficult instruction blocks, $b(i, j)$'s (remember that the higher values of j correspond to the more difficult instruction blocks according to Chapter VII). The slow student would be progressed very slowly through the course.

Present Level (i)	Chosen Block (j)	Behavior (k)*	New Level (V(i, j, k))
1	1	1.00	2
2	3	1.00	4
4	5	1.00	6
6	3	1.00	8 Excellent
8	3	1.00	10 Student
10	3	1.00	11
11	1	1.00	end
1	1	.50	2
2	3	.50	3
3	3	.50	4
4	4	.50	5
5	3	.50	5
5	3	.50	5
5	3	.93	6 Average
6	3	.50	7 Student
7	3	.50	8
8	3	.50	9
9	3	.50	10
10	3	.50	11
11	1	.50	end
1	1	.22	1
1	3	.28	2
2	3	.22	3
3	3	.13	3
3	3	.50	4
4	4	.14	4
4	4	0.0	4
4	4	.50	5
5	3	.15	5 Slow
5	3	.50	5 Student
5	4	.50	6
6	2	.35	6
6	2	.64	7
7	2	.49	8
8	2	.36	9
9	2	.49	9
9	2	.71	10
10	2	.71	11
11	1	.69	end

*U(k) is tabulated in this column---see equation (4.3)

Tutorial Functions Demonstrated for Hypothetical Students

Table 8-2

He would stay at the same level to receive various presentations of the same material until he mastered this material. The slow student would be presented the simple instruction blocks, $b(i, j)$'s. (Again the lower values of j correspond to the simpler instruction blocks per Chapter VII.) This type of tutorial behavior is just what we anticipated the computer-directed machine would exhibit.

Most of the students thought the course was worthwhile. They increased their reading speed an average of 100 words per minute (from 350 to 450 average). While the questions were entirely of the multiple choice variety (for ease in computer grading), several students commented on the generality of the questions. Questions were included asking the student to: pick the best summary of a passage; infer conclusions from the material; recall significant data, etc.

The course is readily changed to reflect new ideas or to correct errors. The course material is all stored on a single magnetic tape which is considerably easier to update than the conventional micro-films associated with teaching machines.

The scope which was used as the tachistoscope was not ideal. Many of the students complained about the legibility of the characters on the scope. Unfortunately the tri-color scope used is inherently inaccurate (see Appendix A). A better tachistoscope would consist of a laboratory oscilloscope with a very low persistence phosphor which was slaved to the Itek flicker free logic (see Appendix A). A simple gating network would be used to specify the high persistence or the low persistence scope.

The text scope proved to be a decidedly attractive output device for teaching machines. Large quantities of text were displayed entirely free from the flicker normally associated with computer displays.

However, the single drawback of this scope was its requirement of large quantities of computer storage to control every motion of the electron beam. The ideal teaching machine output scope would have its own memory as part of a self-contained unit.

In Appendix B the large amount of time necessary to make a decision is discussed. Much of this time is spent in making computer floating-point number calculations. Floating-point arithmetic was necessary because of the large range of numbers involved. The PDP-1 computer used in this experiment does not have an internal floating-point system. Instead floating-point operations are performed by sub-routines. This causes a floating-point operation on the PDP-1 to take about 2000 times longer than such an operation on a machine like the IBM 7094. The computer used for a decision structure such as the one described by this paper should certainly have a built-in floating-point system.

CHAPTER IX

SUGGESTED FUTURE RESEARCH

Many future experiments may be performed with the present teaching machine configuration. Since speed reading is a skill which we all possess to some degree, a pre-test might be given to place the students at an initial level in the course (instead of starting all students off at level 1 block 1 as was done in this experiment). This pre-test could also be used to determine accurately the student's improvement after taking the course.

The value of having a decision structure which is capable of a variable nmax step future search can only be determined after experiments are performed to find the relationship between the decision and the value of nmax. A preliminary study of this relationship was made during this research. The decisions were made for various values of nmax at several levels of the course for the same value of student past history. In all cases the block $b(i, j)$ chosen was the same for all values of nmax tried (see Figure 9-1). However, more experiments will have to be performed to determine this relationship with other student statistics. (In the case described, the statistics were tabulated by the author and were relatively symmetrical between levels.)

Perhaps an analysis should be made considering the value of a large nmax future search versus the cost of a large nmax search. Presumably the value would go up with nmax but the cost would go up with increasing computer time used for making the decision (which goes up faster than the factorial of nmax). Some tradeoff

would then be made of value for cost. Indeed the machine's knowledge of the student might even influence the value of the degree of future search. This value might be less for a new student, about whom the machine has little data, than for a student who is well into the course. This suggests a dynamic criterion for determining nmax based on how far the student is in the course, how accurate the machine's past decisions have been, the cost of an nmax search, etc.

Student's Past History	Level (i)	Chosen Block (j)	nmax
1.0	1	1	1
1.0	1	1	2
.50	1	1	1
.50	1	1	2
.25	1	1	1
.25	1	1	2
1.0	2	3	1
1.0	2	3	2
1.0	2	3	3
.50	2	1	1
.50	2	1	2
.25	2	1	1
.25	2	1	2

Computer-Directed Decisions
versus nmax

Figure 9-1

Often when students were replaced to the same level, they were presented the same block of instruction material (see Table 8-2). This suggests a useful modification to the decision structure which would include a consideration of the number of times the student has been presented each block of instruction material. A choice might first be made between those blocks never presented to the student. If the student has received all of the blocks at a given level, the choice might be made between all blocks given only once, etc. Certainly after the same block is presented more than twice, some alarm device should be set to call in a human teacher. The student would be hung up in an endless loop in this situation which suggests inadequacies in the teaching machine program.

In summary, then, a very powerful teaching machine---the computer-directed teaching machine---has been introduced. Continued experimentation, like Smallwood's miniature geometry course and the present speed reading course, will determine how well this machine teaches in comparison with existing teaching machines.

APPENDIX A

Speed Reading Input-Output Devices

It was necessary to have two devices available for the speed reading course:

1. A tachistoscope.

2. A text display scope.

The computer scopes were chosen to represent both devices. It was not possible to use the same scope, however, for both devices. The tachistoscope requires the use of a scope with a low persistence phosphor. (The image on the tachistoscope must decay within 1/100 second in order to make the 1/100 second duration sweeps possible.) The text display function (on the other hand) requires a steady, flicker-free image. Therefore a scope with a high persistence phosphor is in order here.

The tri-color display scope of the AFCRL PDP-1 dual computer facility was chosen for the tachistoscope because of its phosphors' low persistence. Several tachistoscopic innovations were suggested by the use of a device with an "electronic shutter" as opposed to a mechanical shutter. With most tachistoscopes a mechanical shutter momentarily exposes the text. The shutter movement is generally vertical (spring loaded gravity shutters are generally used). An obvious improvement would be a horizontal exposure of the material. (We read from left to right---horizontally.) This operation is trivial with an electronic shutter.

Cole's[4] suggestion of the initial exaggeration of phrase groupings (by spreading them apart) was extended with the electronic tachistoscope.

The initial material was spread apart not only physically but also chrono-logically. That is, the phrases were sequentially displayed from left to right with pauses between each phrase displayed. The entire phrase was displayed at effectively the same instant because of a high speed, multiple sweep technique. As a control some of the material was presented without this sequential feature (a simple left to right slow speed "single-sweep" shutter was used). The students who were given both types of display sweeps performed better with the phrase type of display. Once again the conditioning stimuli (both the physical and chronological spacing between phrases) were vanished as the student progressed through the tachistoscope portion of the course. The student gradually replaced these stimuli with his own judgements about phrase groupings.

Material was prepared for the tachistoscope portion of the course on a flexowriter with a very simple format.

> e. g.
>
> p100, 25
>
> now / is the time / for all good men

For each segment of material the flash duration time in milliseconds was specified (in this case 100 milliseconds). If the flash duration specification was preceded by a "p", the phrase sweep mode was used. The second number presented is the optional specification of the per-centage of the sweep time to be spent at the delays between phrases (marked by the "/" characters).

The black and white, incremental or line segment display of the AFCRL facility was chosen for the text display both because of its phosphor's high persistence and because of the unique, flicker-free

capabilities engineered into this scope by the Itek Corporation of Lexington, Mass.

Again material was prepared for text display on a flexowriter. The course comments can be intermixed with tachistoscope as well as text material by enclosing the comments within parentheses. Material so enclosed is displayed on the black and white scope until the page turning switch is activated. When the reading rate was to be measured for a passage, the count of the total number of words in the passage must preceed the passage.

e.g.

10

(Now is the time for all good men to vote.)

Again the material is displayed until the page turner is engaged. Questions about text are set off by overbars $"^-"$. The first overbar initiates the question and the second overbar terminates the question. The second overbar is followed by the number corresponding to the correct answer.

e.g.

$\overline{\text{The}}$ product of x times x is:

1) x

2) x^0

3) x^2

4) indefinite

$\overline{3}$

Questions are displayed until first a typewritten answer is given then the page turner is engaged.

APPENDIX B

Decision Structure Computer Techniques

With this project a significant improvement in computer versatility was made over Smallwood's decision structure computer realization. This realization can be programmed in two ways:

1. A different segment of the program may be used for each of the nmax teaching cycles to be scanned, and one program segment would be used to perform the truncation estimation. Here one program segment is necessary for each cycle of the search; if nmax were three, three program segments would be needed.

2. The same segment of the program may be used recursively for each of the nmax teaching cycles scanned, and again a single program segment would be used to perform the estimations at truncation. Now if nmax were three, one program segment plus a push-down structure to implement the recursive nature would be required.

The first approach, used by Smallwood, has the advantage of simple, straight-forward programming. The second approach, used with this project, while difficult to program has the advantage of allowing nmax to be any value without increasing the size of the computer program. Since the programming effort is only performed once for a teaching machine, the second approach seems to be the more rewarding choice.

Because the second approach is difficult to program, a discussion of the programming techniques used in this experiment is presented. Recalling the nmax step truncation maximization equation (4.10).

$$\text{Max}_{j_n} \overline{U}_{i_n j_n}(h_n) = \text{Max}_{j_n} \sum_{k_n} P_{i_n j_n}(h_n | k_n) \cdots \text{Max}_{j_M} \sum_{k_M} P_{i_M j_M}(h_M | k_M) \int_{\text{all } U} U f_{i_\ell}(U | h_\ell) dU$$

and defining:

n - the number of teaching cycles the search is ahead of the student

i_n, j_n, k_n, h_n - the current level, branch, measured behavior range, value of past history---respectively

i_{n+1} - the value of the new level after placement

Umax - the maximum value of the function U which is indicative of learning

$j_n$max - the j_n which yields Umax---the decision

Umaxt, ht - temporary storage for Umax, h respectively

kmax - the maximum number of ranges into which measured behavior can be fitted (5 ranges in this experiment)

jcount - the total number of instruction blocks available at a given level. This is a function called "count" of i_n,

a block diagram of the program segment used in the variable step search computer technique is presented in Figure B-1. The program is made recursive by saving the parameters of the list in a push-down list until it is necessary to recall the parameters by pulling them for the push-down table. The list parameters are: i_n, j_n, k_n, h_n, ht, $j_n$max, U, Umax, jcount.

The decision search process is started at the entry called "present". Here the search begins for the end of the course or the truncation value of n (whichever comes first) whereupon the appropriate value of U is estimated. As every possible path (up to truncation) is considered, the value of n will change, and the parameters that make

each cycle unique will be restored at the proper cycle by the recursive push-pull scheme. When the search reaches to the future, the parameter list is pushed down or saved. When the search retreats to the past, the parameter list is pulled back or restored. Ultimately the search is completed and the decision is made.

Generally speaking the closer the search is to an exhaustive search, the better the decision. That is, the larger the value of nmax, the more reliable the decision. In the present experiment, computer (A) was making the decisions while computer (B) was controlling the input-output equipment. Actually computer (A) was making decisions based on each of the possible ranges into which the student's measured behavior for the current instruction block might fit. First the decisions were made for nmax = 1 since these decisions took the least time. If the student was still receiving the instruction material when decisions had been calculated for all possible behavior ranges, new decisions were made for nmax = 2, and so on. The decision actually used to pick the instruction block for the student was always the best decision currently available.

PRESENT

FUTURE

PUSH (list)

$i_n = i_{n+1}$

$n = n + 1$

$n \leq M$

$n = \ell$

Find estimate of U_{max}

NO

$i_{n+1} = i \ell$?   YES

$U = P_{i_n j_n}(k_n|h_n) \cdot$

$\int U F_{i\ell}(U|h_1) dU$

$n = n-1$

$U = P_{i_n j_n}(k_n|h_n)$

$U_{maxt} + U$

Truncation Estimate

$i_{n+1} = V(i_n, j_n, k_n)$

$h_t = h_n$

$h_{n+1} = W(h_n, k_n)$

$k_n = k_n + 1$

YES

$h_n = h_t$

$k_n \vee k_{max}$?

$U = P_{i_n j_n}(k|h_n)$

$U_{maxt} + U$

$n = n-1$

$U_{maxt} = U_{max}$

PULL (list)

PAST

$u = 0$

$k_n = 1$

$j_n = j_n + 1$

$U_{max} = U$

$j_{nmax} = j_n$ NO

NO

$U \leq U_{max}$?   YES

$j_n \leq jcount$?

$m = 1$?

NO

$U_{max} = 0$

$j_n = 1$

jcount=count(i_n)

Decision is j_{nmax}

YES

DONE

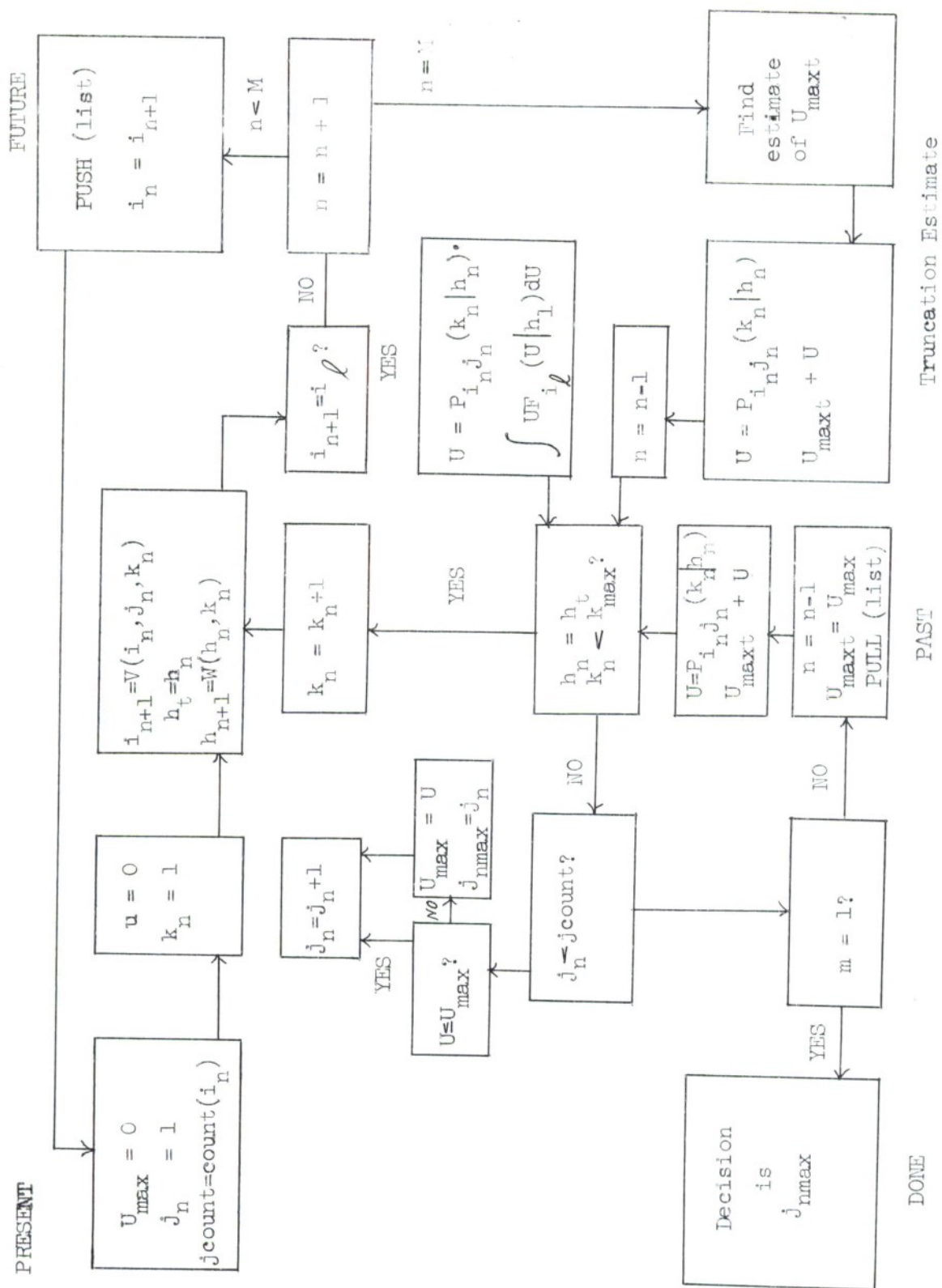Block Diagram of Decision Program

Figure B-1

APPENDIX C

Maximum Likelihood Estimation

The conditional probability density function $g_{ij}(h|k)$ is important in the decision making calculations. It is necessary to estimate a probability function for $g_{ij}(h|k)$ based on the experimentally observed past histories for students who were instructed by block $b(i, j)$ and whose test performance for block $b(i, j)$ fell in range k. That is, a density function must be found to represent the set of observations of history (h) for the total number (N) of students who have taken the course.

Assume the set of values of history $\{h_1, h_2, h_3, \ldots, h_N\}$ have been picked from a Beta function of unknown parameters r and s. Beta functions are assumed because of their generality. For a fixed level (i), branch (j), and test behavior range (k)

$$g_{ij}(h|k) = g(h; r, s)$$

where   (C.1)

$$g(h; r, s) = \frac{h^{r-1}(1-h)^{s-1}}{B(r, s)} \qquad 0 \leq h \leq 1$$

and it is understood that the subscripts i, j, k remain constant, but they are dropped for convenience.

The normalization factor $B(r, s)$ is necessary for the density function to integrate (over h) to unity and is expressed in terms of the gamma function.

$$\frac{1}{B(r, s)} = \frac{\Gamma(r+s)}{\Gamma(r)\,\Gamma(s)}$$

A maximum likelihood estimate[7,18] for the unknown parameters
r and s would require that the value of the N-dimensional joint probabi-
lity density function for the N observations (called the likelihood function,
L) with the appropriate r and s be a maximum. Such an estimation
criterion was used in this research project. If we assume these obser-
vations are independent, the likelihood function of the observations and
the unknown parameters r and s is equal to the product of the values
of the Beta function for each observed history (h)

$$L(h_1, h_2, \ldots, h_N; r, s) = \prod_{m=1}^{N} \frac{h_m^{r-1}(1-h_m)^{s-1}}{B(r,s)} \qquad (C.2)$$

To find the maximum likelihood estimate of r and s in terms
of the experimental data, we must maximize L with respect to r and s.
Maximizing the logarithm of L is equivalent to maximizing L and is
considerably easier. This maximization is performed by separately
taking the partial derivative of the logarithm of the likelihood function
with respect to each unknown parameter (r, s) and setting each partial
derivative equal to zero.

$$\log(L) = -N[\log \Gamma(r) + \log \Gamma(s) - \log \Gamma(r+s)]$$
$$+ (r-1)\sum_{m=1}^{N} h_m + (s-1)\sum_{m=1}^{N} (1-h_m) \qquad (C.3)$$

$$\frac{\partial \log(L)}{\partial r} = -N \frac{\partial \log \Gamma(r)}{\partial r} + N \frac{\partial \log \Gamma(r+s)}{\partial r} + \sum_{m=1}^{N} \log h_m = 0 \qquad (C.4)$$

$$\frac{\partial \log(L)}{\partial s} = -N\frac{\partial \log \Gamma(s)}{\partial s} + N\frac{\partial \log \Gamma(r+s)}{\partial s} + \sum_{m=1}^{N} \log(1-h_m) = 0$$

$$(C.5)$$

The function $\dfrac{\partial \log \Gamma(x)}{\partial x}$ is called the psi function and is tabulated[10].

$$\frac{\partial \log \Gamma(x)}{\partial x} = \psi(x) \qquad (C.6)$$

Recalling

$$\frac{\partial f(x+y)}{\partial x} = \frac{\partial f(x+y)}{\partial (x+y)}$$

equations (C.4) and (C.5) become

$$\psi(r+s) - \psi(r) = -\frac{1}{N}\sum_{m=1}^{N} \log(h_m) \qquad (C.7)$$

$$\psi(r+s) - \psi(s) = -\frac{1}{N}\sum_{m=1}^{N} \log(1-h_m) \qquad (C.8)$$

Equations (C.7) and (C.8) are solved for r and s by an iterative procedure[18] for a given N observations of history $\{h_m\}$. In order to update the values of r and s when a single additional history (call it h_n) is observed after one more student has encountered the point (i, j, k) in the course, the old values of three pertinent parameters N, $\sum_{m=1}^{N} \log(h_m)$ $\sum_{m=1}^{N} \log(1-h_m)$ must be known. Proceeding with update process then involves:

updating N to a new value N'

$$N' = N + 1 \tag{C.9}$$

updating $\displaystyle\sum_{m=1}^{N} \log(h_m)$ to a new value $\displaystyle\sum_{m=1}^{N'} \log(h_m)$

$$\sum_{m=1}^{N'} \log(h_m) = \sum_{m=1}^{N} \log(h_m) + \log(h_n) \tag{C.10}$$

updating $\displaystyle\sum_{m=1}^{N} \log(1-h_m)$ to a new value $\displaystyle\sum_{m=1}^{N'} \log(1-h_m)$

$$\sum_{m=1}^{N'} \log(1-h_m) = \sum_{m=1}^{N} \log(1-h_m) + \log(1-h_n) \tag{C.11}$$

APPENDIX D

Bayesian Estimation

The importance of estimating a probability density function $g_{ij}(h|k)$ from a set of empirical histories $\{h_m\}$ as well as the estimation technique used in this project were discussed in Appendix C. However a different method of estimation called Bayesian estimation will be discussed in this section. This method is more difficult to implement on a computer.

Assume that the conditional probability density function of an observation of history (h) at a particular i, j, k given the parameters r and s is:

$$G(h|r, s) = \frac{h^{r-1}(1-h)^{s-1}}{B(r, s)} \tag{D.1}$$

where the i, j, k parameters are held constant but dropped for convenience throughout this section.

A prior exists for the joint probability density function of r and s---call it $w_0(r, s)$. The subscript on w indicates the number of times the function (w) has been updated---zero times here because it is the prior.
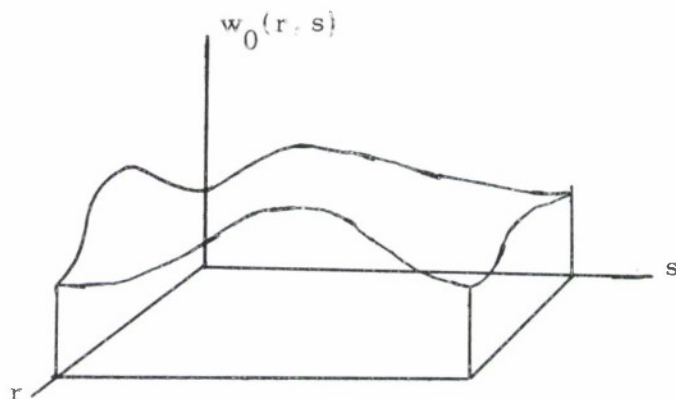
Prior for $w(r, s)$

Figure D1

Then the estimated probability density function values for each history (h) is:

$$\hat{g}_0(h) = \int\int_{\text{all } r+s} w_0(r,s)\, G(h\,|\,r,s)\, dr\, ds \qquad (D.2)$$

The $\wedge$ over a variable means the estimate of that variable.

The subscript on $\hat{g}_0(h)$ indicates the particular function (w) from which the estimated value of $g(h)$ is derived.

When an observation of history (h_1) is made, the conditional probability density function pertinent to the decision calculations is:

$$g_1(h\,|\,h_1)$$

This function is estimated by:

$$\hat{g}_1(h\,|\,h_1) = \int\int_{\text{all } r+s} w_1(r,s\,|\,h_1)\, G(h\,|\,r,s)\, dr\, ds \qquad (D.3)$$

where the function (w_1) is a posterior function based on the prior function (w_0) and the observation h_1. From Bayes' Theorem:

$$w_1(r,s) = \frac{G(h_1\,|\,r,s)\, w_0(r,s)}{\hat{g}_0(h_1)} \qquad (D.4)$$

Using the formula (D.4) a new function (w_1) is generated for use in conjunction with computing the estimate of $g_1(h\,|\,h_1)$ in equation (D.3)

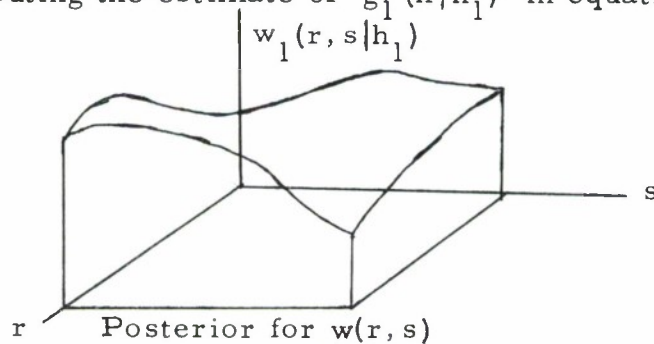$w_1(r,s\,|\,h_1)$

s

r

Posterior for $w(r,s)$

Figure D2

-56-

This suggests a successive technique with which the function (w) may be updated for each new observation of history (h_n) to give the latest estimate of $g_n(h|h_1, \ldots, h_n)$ conditioned upon all of the empirical observations. The formula for successively updating the function (w) is:

$$w_n(r, s|h_1, h_2, \ldots, h_n) = \frac{G(h_n|r, s)w_{n-1}(r, s|h_1, h_2, \ldots, h_{n-1})}{\hat{g}_{n-1}(h_n|h_1, \ldots, h_{n-1})}$$

$$n > 1 \qquad\qquad \text{(D.5)}$$

The estimation of the conditional probability density value for $g(h|h_1, \ldots, h_n)$ becomes:

$$\hat{g}_n(h|h_1, \ldots, h_n) = \int\int_{\text{all } r+s} w_n(r, s|h_1, \ldots, h_n) G(h|r, s) \, dr \, ds$$

$$\text{(D.6)}$$

A realization of this estimation procedure would involve the tabulation of the w function, and updating the w function table. This would involve many double integrations to compute $\hat{g}_{n-1}(h_n|h_1, \ldots, h_{n-1})$. It would also be necessary to perform a double integration every time an estimate of $\hat{g}_n(h|h_1, \ldots, h_n)$ were required unles $\hat{g}_n(h|h_1, \ldots, h_n)$ were also tabulated. This realization would require enormous computer storage and time; hence, the method of Bayesian estimation was considered impractical for a computer-directed decision mechanism.

APPENDIX E

Speed Reading Course Operating Instructions

For those who wish to continue this experiment or to take the speed reading course, the operating instructions are included. Two programs are read into computer (A). First the data base of teaching machine tables is read into core 1 of computer (A), designated as PDP-1c 3, via the paper tape reader. Next the decision structure and control program is read into core 0 of computer (A). The external switch box must be attached to the external sense switch receptacle of computer (B), designated as PDP-1c 4. The teaching machine course tape is threaded onto a tape unit of computer (B). This tape unit is dialed to unit 4. After the paper tape called "RIM Teaching Machine" is read into core 0 of computer (B), the machine is ready to begin the course.

After a student has completed the course, a paper tape summary of the student's path and test behaviors is punched out by computer (A). This paper tape should be checked for rips by placing it in the reader and starting the computer (A) with sense switch 1 up at location 100_8. If there is an error typeout, the computer should be restarted at location 555_8 to get a new punch-out. This paper tape is used to update the teaching machine tables.

The update process may be done after each student or after many students. The update program is read into either computer's core 0. The current teaching machine tables must be located in core 1 of this machine. With all sense switches down, the individual student tapes are threaded into the reader, and the computer is started

at 100_8. When the computer is finished with each tape, it will type "update process completed". At this time either a new tape will be threaded or the computer is restarted at 100_8 with sense switch 2 up for the punch out of the revised teaching machine tables. This punch out may be checked for rips by threading it into the reader, leaving sense switch 2 up, and depressing the read-in switch. If there are any errors, the reader will stop before the end of the tape is reached. The rip free punchout is used as the current teaching machine tables until the update process is performed again.

BIBLIOGRAPHY

1. Baker, William D., "Reading Skills", Prentice-Hall Inc. (1953).

2. Beck, Jacob, "On Some Methods of Programming", 55-62 of Galanter, Ref. 11.

3. Better Reading Program, Encyclopaedia Britannica (1958).

4. Cole, Luella, The Improvement of Reading, Ferrar & Rinehart Inc., (1940).

5. Coulson, J. E. & Silberman, H. F., "Results of an Initial Experiment in Automated Teaching", 452-468 of Lumsdaine and Glaser, Ref. 13.

6. Coulson, J. E., Programmed Learning and Computer Based Instruction, Wiley & Sons (1962).

7. Cramer, Harold, Mathematical Methods of Statistics, Princeton University Press, Princeton, N. J. (1946).

8. Crowder, Norman A., "Automatic Tutoring by Means of Intrinsic Programming", 109-116 of Galanter, Ref. 11.

9. Crowder, Norman A., "Intrinsic and Extrinsic Programming", 58-66, Ref. 6.

10. Davis, H. P., Tables of Higher Mathematical Functions, Principia Press (1933).

11. Galanter, Eugene, Automatic Teaching, The State of the Art, Wiley & Sons, Inc. (1959).

12. Kay, Sylvia C., Reading Critically, Twayne Publishers Inc. (1952).

13. Lumsdaine, A. A., and Glaser, R., Teaching Machines and Programmed Learning: A Source Book, Department of Audio-Visual Instruction, National Educational Association (1960).

14. Lysaught and Williams, A Guide to Programmed Instruction, Wiley & Sons, Inc., (1963).

15. Pask, Gordon, "The Self-Organizing Teacher", 13-18 from Automated Teaching, vol. 1, no. 2.

16. Ryans, David G., "Education Research & Development", 1-8, SDC Magazine, vol. 6, no. 3, (March 1963).

17. Skinner, B. F., and Holland, J. G., "The Use of Teaching Machines in College Instruction", from Lumsdaine and Glaser, Ref. 13.

18. Smallwood, Richard D., A Decision Structure for Teaching Machines, M. I. T. Press (1962).

19. Stolurow, L. M., Teaching by Machine, U. S. Gov. Printing Office OE-34010 (1963).

20. Teachers Conference on the Promotion of Reading, Reading in Action, N. Y. State Publishers, (1957).

21. Wadsworth and Bryand, Introduction to Probability and Random Variables, McGraw-Hill Inc. (1960).

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
M.I.T. Cambridge 39, Massachusetts	UNCLASSIFIED
	2b. GROUP N/A

3. REPORT TITLE

A Generalized Teaching Machine Decision Structure With Application To Speed Reading

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(Last name, first name, initial)*

Strollo, Theodore R.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 64	67	0

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF19(628)2407	ESD-TDR-64-523
b. PROJECT NO. 7682	
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d.	

10. AVAILABILITY/LIMITATION NOTICES

Qualified Requesters May Obtain From DDC.
Aval From AUL.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Decision Sciences Labs., ESD, L.G. Hanscom Field, Bedford, Mass.

13. ABSTRACT

A relatively new type of automated instruction called the "computer-directed" teaching machine is discussed. The computer-directed machine chooses instruction material by making a statistical evaluation of the student's total behavior in cimparison with other student's total behaviors. This machine's statistics are changed as new students take the course. This machine's statistics are changed as new students take the course. The role of this computer-directed machine in the teaching machine field can only be determined after:

1. A technique for comparing teaching machines is developed.

2. More research is performed utilizing the computer-directed machine. In this paper a technique is suggested for comparing teaching machines.

DD FORM 1 JAN 64 **1473**

14.		LINK A		LINK B		LINK C	
KEY WORDS		ROLE	WT	ROLE	WT	ROLE	WT
Teaching Machine Computer Automatic Instruction Speed Reading							

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

_____ ."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

_____ ."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

_____ ."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional